YDY4171 ASSIGNMENT 6

Learning Decision Tree

Results

```
Using measure: random
Average accuracy: 77

Using measure: information_gain
Average accuracy: 93
```

Comments

When using a sample size of 100 I get the average accuracies listed above; Random: 77% and IG: 93%. The reason for this difference lies within how the learning algorithm selects the next node to add to the tree. With information_gain the node is selected depending on entropy, and with random, the node is selected completely random. We see that the accuracy is well above the arbitrary 50% expected with random choice. This means that it is learning, but with the information gain the accuracy is significantly higher.

Attribute decision making algorithms:

```python
if measure == "random":
    # Randomly select an attribute
    return np.random.choice(attributes)
elif measure == "information_gain":
    # Initialize maximum information gain and the best attribute
    max_gain = -np.inf
    best_attribute = None
    for attribute in attributes:
        # Compute the information gain for each attribute
        gain = information_gain(examples, attribute)
        if gain > max_gain:
            max_gain = gain
            best_attribute = attribute
    return best_attribute
```